Internet Appendix for "Dynamic CEO Compensation"

March 4, 2011

B Continuous Time

We now consider the continuous-time analog of the model, assuming $a_t^* = \overline{a}\ \forall\ t$ (from Theorem 4). The CEO's utility is given by:

$$U = \begin{cases} E\left[\int_0^T \rho^t \frac{(c_t h(a_t))^{1-\gamma}-1}{1-\gamma} dt\right] & \text{if } \gamma \neq 1 \\ E\left[\int_0^T \rho^t \left(\ln c_t + \ln h\left(a_t\right)\right) dt\right] & \text{if } \gamma = 1. \end{cases} \tag{56}$$

The firm's returns evolve according to:

$$dR_t = a_t dt + \sigma_t dZ_t$$

where Z_t is a Brownian motion, and the volatility process σ_t is deterministic. We normalize $r_0 = 0$ and the risk premium to zero, i.e. the expected rate of return on the stock is R in each period.

Proposition 3 *(Optimal contract, continuous time, log utility). The continuous-time limit of the optimal contract pays the CEO c_t at each instant, where c_t satisfies:*

$$\ln c_t = \int_0^t \theta_s dR_s + \kappa_t, \tag{57}$$

where θ_s and κ_t are deterministic functions. If short-termism is impossible, the sensitivity θ_t is given by:

$$\theta_t = \begin{cases} \frac{g'(\overline{a})}{\int_t^T \rho^{\tau-s} ds} & \text{for } t \leq L \\ 0 & \text{for } t > L \end{cases}. \tag{58}$$

If short-termism is possible, θ_t is given by:

$$\theta_t = \begin{cases} \frac{\zeta_t}{\int_t^T \rho^{\tau-s} ds} & \text{for } t \leq L + M \\ 0 & \text{for } t > L + M \end{cases}, \tag{59}$$

where:

$$\zeta_s = \begin{cases} \max_{0<i\leq M}\left\{g'(\bar{a}), \frac{q_i}{\rho^i}\zeta_{s-i}\right\} & \text{for } s \leq L \\ \max_{s-L\leq i\leq M}\left\{\frac{q_i}{\rho^i}\zeta_{s-i}\right\} & \text{for } L < s \leq L+M \end{cases}.$$

If private saving is impossible, the constant κ_t is given by

$$\kappa_t = (R + \ln\rho)\, t - \int_0^t \theta_s E\left[dR_s\right] - \int_0^t \frac{\theta_s^2 \sigma_s^2}{2} ds + \underline{\kappa}. \tag{60}$$

If private saving is possible, κ_t is given by

$$\kappa_t = (R + \ln \rho)\, t - \int_0^t \theta_s E\,[dR_s] + \int_0^t \frac{\theta_s^2 \sigma_s^2}{2} ds + \underline{\kappa}. \tag{61}$$

where $\underline{\kappa}$ ensures that the agent is at his reservation utility.

Proposition 4 *(Optimal contract, continuous time, general CRRA utility, with PS constraint). Let σ_t denote the stock volatility. The optimal contract pays the CEO c_t at each instant, where c_t satisfies:*

$$\ln c_t = \int_0^t \theta_s dR_s + \kappa_t, \tag{62}$$

where θ_s and κ_t are deterministic functions. The continuous-time limit of the optimal contract is the following. The sensitivity θ_t is given by:

$$\theta_t = \frac{\rho^t e^{-(1-\gamma)g(\bar{a})} g'(\bar{a})}{\int_t^T \rho^s e^{-(1-\gamma)g(\bar{a})+(1-\gamma)(\kappa_s - \kappa_t)} E_t \left[e^{(1-\gamma)\int_t^s \theta_\tau dR_\tau} \right] ds} \quad \text{for } t \leq L, \tag{63}$$

$$\theta_t = 0 \quad \text{for } t > L.$$

The value of κ_t is:

$$\gamma \kappa_t = (R + \ln \rho) t - (1-\gamma) g(\bar{a}) \mathbf{1}_{t \geq L} - \gamma \int_0^t \theta_s \bar{a} ds + \frac{1}{2} \gamma^2 \int_0^t \theta_s^2 \sigma_s^2 ds + \underline{\kappa}, \tag{64}$$

where $\underline{\kappa}$ ensures that the agent is at his reservation utility.

The implications of the optimal contract are the same as for discrete time, except that the rebalancing of the account is now continuous. As in the discrete time case, the expressions become simpler if $L = T = \infty$. We have

$$\theta = \frac{g'(\bar{a})}{\int_t^\infty \rho^{s-t} e^{k(1-\gamma)(s-t)} e^{(1-\gamma)\theta \bar{a}(s-t) + \frac{1}{2}(1-\gamma)^2 \theta^2 \sigma^2 (s-t)} ds}.$$

Define

$$v(\theta) = \ln \rho + k(1-\gamma) + (1-\gamma)\theta \bar{a} + \frac{1}{2}(1-\gamma)^2 \theta^2 \sigma^2$$

where

$$\gamma k = (R + \ln \rho) t - \gamma \theta \bar{a} + \frac{1}{2} \gamma^2 \theta^2 \sigma^2.$$

We obtain the definition

$$\begin{aligned} v(\theta) &= -\ln \rho + \frac{\gamma - 1}{\gamma}(R + \ln \rho) + \frac{\gamma - 1}{2} \theta^2 \sigma^2 \\ &= \frac{(\gamma - 1) R - \ln \rho}{\gamma} + \frac{\gamma - 1}{2} \theta^2 \sigma^2. \end{aligned}$$

We then have $\theta = \frac{g'(\bar{a})}{\int_t^\infty e^{-v(\theta)(s-t)}ds}$, i.e.

$$\theta = g'(\bar{a})v(\theta).$$

The solution is the one in the discrete time model in the main paper, (27).

C Analysis of Theorem 2

This section provides the analysis behind the comparative statics of the determinants of θ_t, discussed in the main paper shortly after Theorem 2. To study the impact of volatility on the contract, we parameterize the innovations by $\varepsilon_t = \sigma\varepsilon'_t$, where σ indicates volatility. We define the function:

$$G(\theta, \gamma, \sigma) = \frac{\gamma - 1}{\gamma} \ln E\left[e^{-\gamma\theta\sigma\varepsilon'}\right] - \ln E\left[e^{(1-\gamma)\sigma\theta\varepsilon'}\right]$$

in the domain $\theta \geq 0, \sigma \geq 0, \gamma \geq 1$. For instance, when ε' is a standard normal, $G(\theta, \gamma, \sigma) = \theta^2\sigma^2\frac{\gamma-1}{2}$, and G is increasing in θ, γ, and σ.

We also define

$$H(\theta, \gamma, \sigma) = G(\theta, \gamma, \sigma) - \frac{\ln\rho + R}{\gamma}$$

If $\ln\rho + R$ is sufficiently close to 0, then $H(\theta, \gamma, \sigma)$ is increasing in in θ, γ, σ.

Lemma 5 *Consider the domain $\theta \geq 0, \sigma \geq 0, \gamma \geq 1$, in the case where $\phi = 0$, $T = L$ and $a_t^* = a^*$ $\forall t$. Suppose that $H(\theta, \gamma, \sigma)$ is increasing in its arguments in that domain. Then, $\theta_T = g'(a^*)$, and for $t < T$, θ_t is increasing in γ, in σ, and decreasing in ρ. If $H(\theta, \gamma, \sigma)$ is close enough to 0, then θ_t is increasing in t.*

The lemma means that the sensitivity profile is increasing, and becomes flatter as γ and σ are higher. The intuition is thus: a higher γ, a higher σ, or a lower ρ, tend to decrease the relative importance of future consumptions $E\left[\rho^t c_t^{1-\gamma}\right]$. Hence, it is important to give a higher sensitivity to the agent early on. By contrast, when γ is low, future consumptions are more important and so it is sufficient to give a lower sensitivity early on.

Proof Using Theorem 2, simple calculations show, for $t \leq L$,

$$\begin{aligned}
\theta_t &= \frac{g'(a^*)}{\sum_{s=t}^T \rho^{s-t} \prod_{n=t+1}^s e^{-G(\theta_n,\gamma,\sigma)+\frac{1-\gamma}{\gamma}(R+\ln\rho)}} \\
&= \frac{g'(a^*)}{\sum_{s=t}^T \prod_{n=t+1}^s e^{-G(\theta_n,\gamma,\sigma)+\frac{1-\gamma}{\gamma}R+\frac{1}{\gamma}\ln\rho}} \\
\theta_t &= \frac{g'(a^*)}{\sum_{s=t}^T e^{-\sum_{n=t+1}^s (H(\theta_n,\gamma,\sigma)+R)}} \qquad (65)
\end{aligned}$$

We have $\theta_T = g'(a^*)$. Proceeding by backward induction on t, starting at $t = T$, we see that θ_t is increasing in γ: this is because a higher γ increases $H(\theta_n, \gamma, \sigma)$ via the direct effect on H, and the effect on the future θ_n $(n > t)$, so it increases θ_t. The same reasoning holds for the comparative statics with respect to σ and ρ.

The last part of Lemma 5 comes from the fact that when $H \to 0$, $\theta_t \to \frac{g'(a^*)}{\sum_{s=t}^{T} e^{-R(t-s)}}$, which is increasing in t. ■

Another tractable case is the infinite horizon limit, where $T = L \to \infty$. Since the problem is stationary, θ_t is equal to a limit θ. From (65), this satisfies:

$$\theta = g'(a^*)\left(1 - e^{-H(\theta,\gamma,\sigma)-R}\right).$$

For instance, in the continuous-time, Gaussian noise limit,

$$\theta = g'(a^*)\left[\theta^2\sigma^2\frac{\gamma-1}{2} - \frac{\ln\rho + R}{\gamma} + R\right].$$

which gives the solution (27). The sensitivity of incentives (θ) is higher when the agent is more risk-averse (higher γ, provided $\ln\rho + R$ is close enough to 0), there is more risk (higher σ), and the agent is less patient (lower ρ).

D Variable Cost of Effort

This section extends the core model to allow a deterministically varying marginal cost of effort. In practice, this occurs if either the cost function or high effort level changes over time. For example, for a start-up firm, the CEO can undertake many actions to improve firm value (augmenting the boundary effort level) and effort is relatively productive (reducing the cost of effort). However, the scope and productivity of effort declines as the firm matures.

We now allow for a time-varying boundary effort level $\overline{a}_t$ and cost of effort $g_t(\cdot)$. The sensitivity of the contract in Theorem 1 and Proposition 2 (equations (9) and (33)) now becomes:

$$\theta_t = \begin{cases} \frac{g_t'(\bar{a}_t)}{1+\rho+\ldots\rho^{T-t}} \text{ for } t \leq L \\ 0 \text{ for } t > L \end{cases}, \tag{66}$$

if myopia is impossible, and if myopia is possible

$$\theta_t = \begin{cases} \frac{\zeta_t}{1+\rho+\ldots\rho^{T-t}} \text{ for } t \leq L+M \\ 0 \text{ for } t > L+M \end{cases}, \qquad . \tag{67}$$

where

$$\zeta_s = \begin{cases} \max_{1\leq i\leq M}\left\{g_s'(\bar{a}_s), \frac{q_i}{\rho^i}\zeta_{s-i}\right\} \text{ for } s \leq L \\ \max_{s-L\leq i\leq M}\left\{\frac{q_i}{\rho^i}\zeta_{s-i}\right\} \text{ for } L < s \leq L+M \end{cases}.$$

With a non-constant marginal cost of effort, the contract sensitivity θ_t is time-varying, even in an infinite-horizon model. In particular, θ_t is high in the periods in which $g_t'(\bar{a}_t)$ is high. Let $s \leq L$ denote the period in which $g_t'(\bar{a}_t)$ is highest. Even if there is no discounting ($\rho^i = 1$), the CEO may have an incentive to increase r_s at the expense of the signal in period j (where $j \leq s + M$), if the difference in slopes θ_s and θ_j is sufficient to outweigh the inefficiency of earnings inflation ($q_i < 1$). Thus, the sensitivity θ_j will have to rise to be sufficiently close to θ_s to deter such myopia. However, this in turn has a knock-on effect: since θ_j has now risen, the CEO may have an incentive to increase r_j at the expense of r_k (where $k \leq j + M$) and so on. Therefore, if q_i is sufficiently high (to make myopia attractive), the high sensitivity at s forces upward the sensitivity in all periods $t \leq L + M$, even those more than M periods away from s, owing to the knock-on effects. This "resonance" explains the recursive formulation in equation (67), where a high $g_t'(\bar{a}_t)$ may affect the sensitivity for all $t \leq L + M$.

This dependence can be illustrated in a numerical example. We set $T = 5$, $L = 3$, $\rho = 1$, $g_1'(\bar{a}_1) = 2$ and $g_2'(\bar{a}_2) = g_3'(\bar{a}_3) = 1$. If myopia is impossible, the optimal contract is

$$\begin{aligned}
\ln c_1 &= \frac{2}{5}r_1 + \kappa_1 \\
\ln c_2 &= \frac{2}{5}r_1 + \frac{r_2}{4} + \kappa_2 \\
\ln c_3 &= \frac{2}{5}r_1 + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_3 \\
\ln c_4 &= \frac{2}{5}r_1 + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_4 \\
\ln c_5 &= \frac{2}{5}r_1 + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_5.
\end{aligned}$$

Since the marginal cost of effort is high at $t = 1$, the contract sensitivity must be high at $t = 1$ to satisfy the EF condition. However, this now gives the CEO incentives to engage in myopia if it were possible. Assume $M = 1$ and $q_1 > \frac{1}{\sqrt{2}}$. If he engages in myopia that increases r_1 by q units and reduces r_2 by 1 unit, lifetime consumption rises by $2q_1$ units from the former and falls by 1 unit from the latter. Therefore, the sensitivity of the contract at $t = 2$ must increase to remove these incentives. The sensitivity is now $\frac{q_1}{2}$ per period to give a total lifetime reward of $2q_1$. This increased sensitivity at $t = 2$ in turn augments the required sensitivity at $t = 3$, else the CEO would inflate r_2 at the expense of r_3: θ_3 now becomes $\frac{2q_1^2}{3} > \frac{1}{3}$. Therefore, even though the maximum release lag M is 1 and so the CEO cannot take any actions to inflate r_1 at the expense of r_3, the high sensitivity at r_1 still affects the sensitivity at r_3 by changing the sensitivity at r_2. Finally, the contract must remain sensitive to firm returns beyond retirement, to deter the CEO from inflating r_3 at the expense of r_4. The new contract is given by:

$$\begin{aligned}
\ln c_1 &= \frac{2}{5}r_1 + \kappa_1 \\
\ln c_2 &= \frac{2}{5}r_1 + \frac{q_1}{2}r_2 + \kappa_2 \\
\ln c_3 &= \frac{2}{5}r_1 + \frac{q_1}{2}r_2 + \frac{2q_1^2}{3}r_3 + \kappa_3 \\
\ln c_4 &= \frac{2}{5}r_1 + \frac{q_1}{2}r_2 + \frac{2q_1^2}{3}r_3 + q_1^3 r_4 + \kappa_4 \\
\ln c_5 &= \frac{2}{5}r_1 + \frac{q_1}{2}r_2 + \frac{2q_1^2}{3}r_3 + q_1^3 r_4 + \kappa_5.
\end{aligned}$$

This result contrasts with the example in Section 5.2.1 where the possibility of myopia did not change the contract for $t \le L$ under no discounting and a constant marginal cost of effort.

E Additional Proofs

This section contains proofs of lemmas, corollaries and other claims in the main paper.

E.1 Proof of Corollary 1

As $L = T = \infty$ so that we have a constant $\theta_s = \theta$ and $k_s = k$. For notational simplicity we normalize (without loss of generality) $\underline{u} = 0$ and $\bar{a} = 0$. The expected cost of the contract is:

$$\begin{aligned}
\mathcal{C} &= E\left[\sum_{t=1}^{\infty} e^{-Rt} c_t\right] = \sum_{t=1}^{\infty} E\left[\exp\left(-Rt + \ln c_0 + \sum_{s=1}^{t}\theta_s r_s + \sum_{s=1}^{t} k_s\right)\right] \\
&= \sum_{t=1}^{\infty} \exp\left(\left(k - R + \ln E\left[e^{\theta\eta}\right]\right)t + \ln c_0\right) = c_0 \frac{e^{k-R+\ln E\left[e^{\theta\eta}\right]}}{1 - e^{k-R+\ln E\left[e^{\theta\eta}\right]}}
\end{aligned}$$

The value of c_0 is pinned down by the participation constraint:

$$\begin{aligned}
0 = \underline{u} &= E\left[\sum_{t=1}^{\infty}\rho^t \ln c_t\right] = \sum_{t=1}^{\infty}\rho^t\left[\ln c_0 + \sum_{s=1}^{t}\theta_s \bar{a} + \sum_{s=1}^{t} k_s\right] = \sum_{t=1}^{\infty}\rho^t\left[\ln c_0 + kt\right] \\
&= \frac{\rho}{1-\rho}\ln c_0 + \frac{\rho}{(1-\rho)^2}k
\end{aligned}$$

so that: $\ln c_0 = -\frac{1}{1-\rho}k$. Hence

$$\mathcal{C} = e^{-\frac{1}{1-\rho}k}\frac{e^{k-R+\ln E\left[e^{\theta\eta}\right]}}{1 - e^{k-R+\ln E\left[e^{\theta\eta}\right]}}.$$

For the contract without PS, we have $k = R + \ln \rho - \ln E\left[e^{\theta\eta}\right]$, so

$$\mathcal{C}^{NPS} = e^{-\frac{1}{1-\rho}\left(R+\ln\rho-\ln E\left[e^{\theta\eta}\right]\right)}\frac{\rho}{1-\rho}.$$

For the contract with PS, we have $k = R + \ln \rho + \ln E\left[e^{-\theta\eta}\right]$, so

$$\mathcal{C}^{PS} = e^{-\frac{1}{1-\rho}\left(R+\ln\rho+\ln E\left[e^{-\theta\eta}\right]\right)}\frac{\rho e^{\ln E\left[e^{-\theta\eta}\right]+\ln E\left[e^{\theta\eta}\right]}}{1-\rho e^{\ln E\left[e^{-\theta\eta}\right]+\ln E\left[e^{\theta\eta}\right]}}.$$

Thus,

$$\begin{aligned}\Lambda = \frac{\mathcal{C}^{PS}}{\mathcal{C}^{NPS}} &= \frac{(1-\rho)\,e^{-\frac{\rho}{1-\rho}\left(\ln E\left[e^{-\theta\eta}\right]+\ln E\left[e^{+\theta\eta}\right]\right)}}{1-\rho e^{\ln E\left[e^{-\theta\eta}\right]+\ln E\left[e^{\theta\eta}\right]}}\\ &= \frac{1-\rho}{1-\rho e^{\Theta^2\sigma^2}}e^{-\frac{\rho\Theta^2\sigma^2}{1-\rho}}\end{aligned}$$

In the limit of small time intervals, $\ln E\left[e^{-\theta\eta}\right] + \ln E\left[e^{+\theta\eta}\right] \sim \theta^2\sigma^2$, and $1-\rho = \delta$ are small (proportional to the time interval Δt), and $\theta \sim g'(\overline{a})\,\delta$, so

$$\Lambda \sim \frac{\delta e^{-\rho\sigma^2\theta^2/(1-\rho)}}{1-(1-\delta)\left(1+\theta^2\sigma^2\right)} \sim \frac{\delta e^{-\sigma^2\theta^2/(1-\rho)}}{\delta-\theta^2\sigma^2} = \frac{e^{-\sigma^2\theta^2/(1-\rho)}}{1-\frac{\theta^2\sigma^2}{1-\rho}}.$$

E.2 Proof of Theorem 4

We wish to show that, if baseline firm size X is sufficiently large, the optimal contract implements high effort ($a_t \equiv \overline{a}$ for all t).

Fix any contract (A, Y) that is incentive compatible and gives expected utility $\underline{u}$, where $A = (a_1, ..., a_L)$ is the effort schedule, $a_t : [\underline{\eta}, \overline{\eta}]^t \to [0, \overline{a}]$, and $Y = (y_1, ..., y_T)$ is the payoff schedule, $y_t : [\underline{\eta}, \overline{\eta}]^t \to \mathbb{R}$. The timing in each period is as follows: the agent reports noise η_t, then is supposed to exert effort $a_t(\eta_1, ..., \eta_t)$. If the return is $\eta_t + a_t(\eta_1, ..., \eta_t)$ he receives payoff $y_t(\eta_1, ..., \eta_t)$, else he receives a payoff that is sufficiently low to deter such "off-equilibrium" deviations. We require this richer framework, since in general the noises might not be identifiable from observed returns (when $\eta_t + a_t(\eta_1, ..., \eta_t) = \eta'_t + a_t(\eta_1, ..., \eta_{t-1}, \eta'_t)$ for $\eta_t \neq \eta'_t$). Note that the required low payoff may be negative. A limited liability constraint would be simple to address, e.g. by imposing a lower bound on $\underline{\eta}$. We will denote $(\eta_1, ..., \eta_t)$ by $\boldsymbol{\eta}_t$.

To establish the result it is sufficient to show that we can find a different contract (A^*, Y^*) that implements high effort ($a_t \equiv \overline{a}$ for all t), and is not significantly costlier than (A, Y), in the sense that

$$E\left[\sum_{t=1}^{T} e^{-rt}(y_t^*(\boldsymbol{\eta}_t) - y_t(\boldsymbol{\eta}_t))\right] \leq h(E\left[\overline{a} - a_1(\eta_1)\right], ..., E\left[\overline{a} - a_L(\boldsymbol{\eta}_L)\right]), \tag{68}$$

for some linear function h, $h : \mathbb{R}^L \to \mathbb{R}$, with $h(0,...,0) = 0$. This is sufficient, because if initial firm size X is sufficiently large, then for every sequence of noises and actions, firm value $Xe^{\sum_{s=1}^{t-1}(\eta_s+a_s(\boldsymbol{\eta}_s))+\underline{\eta}}$ is greater than D, where D is the highest sensitivity coefficient of h. This in turn implies

$$Xe^{\sum_{s=1}^{t-1}(\eta_s+a_s(\boldsymbol{\eta}_s))+\underline{\eta}} \times E\left[e^{\overline{a}} - e^{a_t(\boldsymbol{\eta}_t)}\right] \geq D \times E\left[\overline{a} - a_t(\boldsymbol{\eta}_t)\right], \tag{69}$$

and so the benefits of implementing high effort outweigh the costs, i.e. the RHS of (68) exceeds the LHS of (68). To keep the proof concise we assume $\rho e^r = 1$, $T = L$ and the noises η_t are independent across time. The general case is proven along analogously.

We introduce the following notation. For any contract (A, Y) and history $\boldsymbol{\eta}_t$ let $u_t(\boldsymbol{\eta}_t) = \frac{[y_t(\eta_t)e^{-g(a_t(\eta_t))}]^{1-\gamma}}{1-\gamma}$ (or $u_t(\boldsymbol{\eta}_t) = \ln y_t(\boldsymbol{\eta}_t) - g(a_t(\eta_t))$ for $\gamma = 1$) denote the CEO's stage game utility for truthful reporting in period t after history $\boldsymbol{\eta}_t$ when he consumes his income, let $U_t(\boldsymbol{\eta}_t) = E_t\left[\sum_{s=t}^{L}\rho^{s-t}u_s(\boldsymbol{\eta}_s)\right]$ denote his continuation utility, and $mu_t(\boldsymbol{\eta}_t) = y_t^{-\gamma}(\boldsymbol{\eta}_t)e^{-(1-\gamma)g(a_t(\boldsymbol{\eta}_t))}$ denote his marginal utility of consumption. We divide the proof into the following six steps.

Step 1. Local necessary conditions. First, we generalize the local effort constraint (5) to contracts that need not implement high effort.

Lemma 6 *Fix an incentive compatible contract (A, Y), with each $a_t(\boldsymbol{\eta}_{t-1}, \cdot)$ continuous almost everywhere and bounded on every compact subinterval, and a history $\boldsymbol{\eta}_{t-1}$. The CEO's continuation utility $U_t(\boldsymbol{\eta}_{t-1}, \eta_t)$ must satisfy the following:*

$$U_t(\boldsymbol{\eta}_{t-1}, \eta_t) = U_t(\boldsymbol{\eta}_{t-1}, \underline{\eta}) + \int_{\underline{\eta}}^{\eta_t}[y_t(\boldsymbol{\eta}_{t-1}, x)e^{-g(a_t(\boldsymbol{\eta}_{t-1},x))}]^{1-\gamma}g'(a_t(\boldsymbol{\eta}_{t-1}, x))dx, \tag{70}$$

with $y_t(\boldsymbol{\eta}_t) > 0$.

Step 2. Bound on the cost of incentives per period. For any history $\boldsymbol{\eta}_{t-1}$ and contract (A, Y), consider "repairing" the contract at time t as follows. Following any history $\boldsymbol{\eta}_{t-1}, \eta$, multiply all the payoffs by the appropriate constant $\zeta(\boldsymbol{\eta}_{t-1}, \eta)$ such that the continuation utilities $U_t^{\#}(\boldsymbol{\eta}_{t-1}, \eta_t)$ for the resulting contract satisfy (70) with $a_t(\boldsymbol{\eta}_{t-1}, \eta_t) = \overline{a}$ for all η_t. In other words, the local EF constraint for high effort at time t after history $\boldsymbol{\eta}_{t-1}$ is satisfied. The following Lemma bounds the expectation of how much we have to scale up the payoffs by the expectation of how much the target effort falls short of the boundary effort level.

Lemma 7 *Fix an incentive compatible contract (A, Y) and a history $\boldsymbol{\eta}_{t-1}$, and consider the contract $(A^{\#}, Y^{\#})$ such that:*

$$a_t^{\#}(\boldsymbol{\eta}_{t-1}, \eta_t) = \overline{a} \text{ for all } \eta_t, \text{ else } a_s^{\#} \equiv a_s,$$
$$y_s^{\#}(\boldsymbol{\eta}_s) = y_s(\boldsymbol{\eta}_s) \times \zeta(\boldsymbol{\eta}_{t-1}, \eta_t) \text{ if } \boldsymbol{\eta}_{s|t} = \eta_{t-1}, \eta_t, \text{ and else } y_s^{\#}(\boldsymbol{\eta}_s) \equiv y_s(\boldsymbol{\eta}_s),$$

where $\zeta(\boldsymbol{\eta}_{t-1}, \eta_t) \geq 1$ *is the unique number such that* $U_t^{\#}(\boldsymbol{\eta}_{t-1}, \underline{\eta}) = U_t(\eta_{t-1}, \underline{\eta})$ *and*

$$U_t^{\#}(\boldsymbol{\eta}_{t-1}, \eta_t) = U_t^{\#}(\boldsymbol{\eta}_{t-1}, \underline{\eta}) + \int_{\underline{\eta}}^{\eta_t} [\zeta(\boldsymbol{\eta}_{t-1}, x) y_t(\boldsymbol{\eta}_{t-1}, x) e^{-g(\overline{a})}]^{1-\gamma} g'(\overline{a}) dx. \tag{71}$$

Then:

$$E_{t-1}\left[\zeta(\boldsymbol{\eta}_{t-1}, \eta_t)\right] \leq \varphi(E_{t-1}\left[\overline{a} - a_t(\boldsymbol{\eta}_t)\right]), \tag{72}$$

where $\varphi(x) = e^{g'(\overline{a}) \sup \frac{g''}{fg'^2} x}\left(1 + \mathbf{1}_{\gamma<1} e^{g(\overline{a})-g(\underline{a})} g'(\overline{a})(1-\gamma)x\right)$ *for* $\gamma \neq 1$,
$\varphi(x) = e^{g'(\overline{a}) \sup \frac{g''}{fg'^2} x}\left(1 + e^{g(\overline{a})-g(\underline{a})} g'(\overline{a}) x\right)$ *for* $\gamma = 1$, *and* f *is the pdf of noise* η.

Step 3. Constructing the contract that satisfies the local EF constraint in every period. We want to use the procedure from step 2 to construct a new contract (A^x, Y^x) that implements high effort, satisfies the local EF in every period, and has a cost difference over (A, Y) that is bounded by how much (A, Y) falls short of the contract that implements high effort. For this we need the following Lemma.

Lemma 8 *For a contract* (A, Y) *and any* $\zeta > 0$ *consider the contract* $(A, \zeta Y)$ *in which all the payoffs are multiplied by* ζ,

i) if (A, Y) *satisfies the local EF constraint then so does* $(A, \zeta Y)$*;*

ii) if (A, Y) *satisfies the local PS constraint then so does* $(A, \zeta Y)$.

Given an incentive compatible contract (A, Y), we construct the contract (A^x, Y^x) as follows. The contract always prescribes high effort. Regarding the payoffs, for any period t after a history $\boldsymbol{\eta}_{t-1}$ we first multiply all payoffs after history (η_{t-1}, η) with fixed constants $\zeta(\boldsymbol{\eta}_{t-1}, \eta) > 1$ as in Lemma 7 so that the resulting utilities $U_t^{\#}(\boldsymbol{\eta}_t)$ satisfy (71). Then we multiply all payoffs following history $\boldsymbol{\eta}_{t-1}$ by the appropriate constant $\zeta^{pu}(\boldsymbol{\eta}_{t-1}) < 1$ so that for the resulting contract (A^x, Y^x) we obtain the original promised utility, i.e. $U_{t-1}(\boldsymbol{\eta}_{t-1}) = U_{t-1}^x(\boldsymbol{\eta}_{t-1})$. By construction and the above Lemmas, the contract (A^x, Y^x) satisfies the local EF constraint. In particular, due to Lemma 8, repairing the contract after history $\boldsymbol{\eta}_{t-1}$ will not upset the local EF constraint after history $(\boldsymbol{\eta}_{t-1}; \eta_t)$.

The original contract (A, Y) satisfies the local PS constraint, i.e. the current marginal utility of consumption always equals the next-period expected marginal utility. Providing incentives for high effort in contract (A^x, Y^x) upsets this condition. In the following two steps, given (A^x, Y^x), we construct the contract (A^*, Y^*) that also satisfies the local PS constraint and is not much costlier. In particular, we show that the extent to which the marginal utilities of consumption in (A^*, Y^*) depart from the marginal utilities in (A^x, Y^x) is bounded by the extent to which effort falls short of the high effort level in contract (A, Y).

Step 4. Bound on the decrease of expected MU of consumption per period. We split this step into two Lemmas. The first bounds the expected decrease in marginal utility of consumption from providing incentives for high effort in the current period, as in step 2. The second bounds the decrease in expected marginal utility by the expected decrease of the marginal utility.

Lemma 9 *Fix any history* $\boldsymbol{\eta}_{t-1}$ *and look at the original contract* (A, Y) *and the contract* $(A^{\#}, Y^{\#})$ *from step 1. Then:*

$$E_{t-1}\left[\frac{mu_t^{\#}(\boldsymbol{\eta}_{t-1}, \eta_t)}{mu_t(\boldsymbol{\eta}_{t-1}, \eta_t)}\right] \geq$$
$$e^{-\gamma g'(\overline{a}) \sup \frac{g''}{fg'^2} E_{t-1}[\overline{a} - a_t(\boldsymbol{\eta}_t)]} \left(1 - \mathbf{1}_{\gamma<1} e^{-(1+\gamma)(1-\gamma)[g(\overline{a})-g(\underline{a})]} g'(\underline{a})(1-\gamma)(1+\gamma) E_{t-1}\left[\overline{a} - a_t(\boldsymbol{\eta}_t)\right]\right).$$

Lemma 10 *Fix any history* $\boldsymbol{\eta}_{t-1}$ *and look at any two contracts* (A^l, Y^l) (A^h, Y^h) *with positive payoffs that satisfy (70) and for every* η_t, $mu_t^l(\boldsymbol{\eta}_{t-1}, \eta_t) \leq mu_t^h(\boldsymbol{\eta}_{t-1}, \eta_t)$. *Then, for some* $D_2 > 0$:

$$\frac{E_{t-1}\left[mu_t^l(\boldsymbol{\eta}_{t-1}, \eta_t)\right]}{E_{t-1}\left[mu_t^h(\boldsymbol{\eta}_{t-1}, \eta_t)\right]} \geq 1 - D_2\left(1 - E_{t-1}\left[\frac{mu_t^l(\boldsymbol{\eta}_{t-1}, \eta_t)}{mu_t^h(\boldsymbol{\eta}_{t-1}, \eta_t)}\right]\right).$$

Step 5. Constructing the contract that satisfies the local PS constraint in every period. Providing incentives for high effort in (A^x, Y^x) at (say) time L affects the marginal utility of consumption in period L and upsets the PS constraint in period $L-1$. However, restoring the PS constraint in period $L-1$ will affect the marginal utility of consumption in period $L-1$ and so upset the PS constraint in period $L-2$, and so on. In the following Lemma we bound this overall effect using Lemma 9 and iteratively Lemma 10.

Lemma 11 *There is a contract* (A^*, Y^*) *that implements maximal effort and satisfies the local EF and PS constraints, and for every history* $\boldsymbol{\eta}_t$*:*

$$\frac{mu_t^*(\boldsymbol{\eta}_t)}{mu_t^x(\boldsymbol{\eta}_t)} \geq \prod_{s=t+1}^{L} \phi^{s-t}(E_t\left[\psi\left(E_{s-1}\left[\overline{a} - a_s(\boldsymbol{\eta}_s)\right]\right)\right]), \tag{73}$$

where $\phi(x) = 1 - D_2(1-x)$, $\psi(x) = e^{-\gamma g'(\overline{a}) \sup \frac{g''}{fg'^2} x}\left(1 - \mathbf{1}_{\gamma<1} e^{-(1+\gamma)(1-\gamma)[g(\overline{a})-g(\underline{a})]} g'(\underline{a})(1-\gamma)(1+\gamma)x\right)$.

Step 6. Bounding the cost difference (68). By construction, contract (A^*, Y^*) from Lemma 11 implements high effort, causes the local EF constraint to bind, satisfies the local PS constraint and leaves the CEO with the expected discounted utility $\underline{u}$. Therefore it is identical to the contract from Theorem 2, and so also satisfies the global constraints (Theorem 3). It therefore remains to prove (68).

One can verify that for some $D_3 > 0$ for every history $\boldsymbol{\eta}_t$ we have $y_t^*(\boldsymbol{\eta}_t) < D_3$. Moreover, for any $a, b, c \in \mathbb{R}$,

$$a - b \leq a\left(\max\{\frac{a-c}{c}, 0\} + \max\{\frac{c-b}{b}, 0\}\right) = a\left(\max\{\frac{a}{c}, 1\} - 1 + \max\{\frac{c}{b}, 1\} - 1\right).$$

Consequently,

$$E\left[\sum_{t=1}^{L} e^{-rt}(y_t^*(\boldsymbol{\eta}_t) - y_t(\boldsymbol{\eta}_t))\right] \leq D_3 \times E\left[\sum_{t=1}^{L} e^{-rt}\left(\max\{\frac{y_t^*(\boldsymbol{\eta}_t)}{y_t^x(\boldsymbol{\eta}_t)}, 1\} - 1 + \max\{\frac{y_t^x(\boldsymbol{\eta}_t)}{y_t(\boldsymbol{\eta}_t)}, 1\} - 1\right)\right] \leq$$

$$\leq D_3 \times E\left[\sum_{t=1}^{L} e^{-rt}\left(\left(\prod_{s=t+1}^{L} \phi^{s-t}(E_t\left[\psi\left(E_{s-1}\left[\overline{a} - a_s(\boldsymbol{\eta}_s)\right]\right)\right])\right)^{-\frac{1}{\gamma}} - 1 + \varphi\left(E_{t-1}\left[\overline{a} - a_t(\boldsymbol{\eta}_t)\right]\right) - 1\right)\right],$$

where φ is as in Lemma 7, while ϕ and ψ are as in Lemma 11. All functions φ, ϕ, ψ, $\prod_{s=t+1}^{L} x_s$ and $x^{-\frac{1}{\gamma}}$ are continuously differentiable and take value 1 for argument(s) equal to 1, whereas $\overline{a} - a_t(\boldsymbol{\eta}_t)$ is bounded. Therefore there is a linear function $h : \mathbb{R}^L \to \mathbb{R}$ with $h(0, ..., 0) = 0$ such that (68) is satisfied.

The above proof is for the case where private saving is possible as this is the more complex case. If $\gamma = 1$ and private saving is impossible, step 4 is not needed and Lemma 11 in step 5 and step 6 become significantly simpler.

E.3 Contract with CARA Utility and Additive Preferences

With these preferences, the agent has period utility

$$u(c, a) = -e^{-\gamma(c - g(a))}.$$

The derivation of the local constraints and the contract are analogous to the paper. Consider a two period model with no discounting. From EF we have:

$$c_2(r_1, r_2) = B(r_1) + g'(\overline{a}) \times r_2.$$

PS yields:

$$\begin{aligned}
\frac{\partial U}{\partial c_1} &= E_1\left[\frac{\partial U}{\partial c_2}\right] \\
e^{-\gamma(c_1 - g(\overline{a}))} &= E_1\left[e^{-\gamma(B(r_1) + g'(\overline{a}) \times r_2 - g(\overline{a}))}\right], \\
c_1 - g(\overline{a}) &= B(r_1) - g(\overline{a}) - \frac{\log E_1\left[e^{-\gamma(g'(\overline{a}) \times r_2)}\right]}{\gamma}, \\
c_1(r_1) &= B(r_1) + k.
\end{aligned}$$

and so we have

$$\begin{aligned}
c_1(r_1) &= \theta_1 r_1 + k_1, \\
c_2(r_1, r_2) &= \theta_1 r_1 + \theta_2 r_2 + k_1 + k_2,
\end{aligned}$$

similar to the main paper.

E.4 Negative Effect of Short-Termism

We show that the condition (31) is sufficient for myopia to have a negative impact on the expected terminal dividend. Fix the effort strategy to be the high effort strategy. Consider any time t and assume that it has been shown that any myopic actions past time t are suboptimal. We must establish that:

$$e^{\sum_{i=1}^{M}\lambda_i(E[m_{t,i}(\eta_{t+i})])}E\left[e^{\sum_{s=t+1}^{M}\left[\eta_s-\left(\sum_{r=s-M}^{t}\right)m_{r,s-r}(\eta_s)\right]}\right]\leq E\left[e^{\sum_{s=t+1}^{M}\left[\eta_s-\left(\sum_{r=s-M}^{t-1}\right)m_{r,s-r}(\eta_s)\right]}\right].$$

For any $i\leq M$ we have:

$$\begin{aligned}
&e^{\lambda_i(E[m_{t,i}(\eta_{t+i})])}E\left[e^{\eta_{t+i}-\left(\sum_{r=t+i-M}^{t}\right)m_{r,t+i-r}(\eta_{t+i})}\right]\\
&\leq e^{\lambda_i(E[m_{t,i}(\eta_{t+i})])}E\left[e^{\eta_{t+i}-\left(\sum_{r=t+i-M}^{t-1}\right)m_{r,t+i-r}(\eta_{t+i})}-e^{\underline{\eta}-M\times\overline{m}}m_{t,i}\left(\eta_{t+i}\right)\right]\leq\\
&\leq e^{\lambda_i(E[m_{t,i}(\eta_{t+i})])}E\left[e^{\eta_{t+i}-\left(\sum_{r=t+i-M}^{t-1}\right)m_{r,t+i-r}(\eta_{t+i})}\right]\left(1-\frac{e^{\underline{\eta}-M\times\overline{m}}}{E\left[e^{\eta}\right]}E\left[m_{t,i}\left(\eta_{t+i}\right)\right]\right)\leq\\
&\leq e^{\lambda_i(E[m_{t,i}(\eta_{t+i})])}E\left[e^{\eta_{t+i}-\left(\sum_{r=t+i-M}^{t-1}\right)m_{r,t+i-r}(\eta_{t+i})}\right]e^{-\frac{e^{\underline{\eta}-M\times\overline{m}}}{E[e^{\eta}]}E[m_{t,i}(\eta_{t+i})]}\leq\\
&\leq E\left[e^{\eta_{t+i}-\left(\sum_{r=t+i-M}^{t-1}\right)m_{r,t+i-r}(\eta_{t+i})}\right],
\end{aligned}$$

where the first inequality follows from the Mean Value Theorem.

E.5 Proofs of Lemmas

Proof of Lemma 2 Let

$$P_s((b_t)_{t\leq T})=e^{\sum_{n=1}^{s}j_n(b_n)},$$
$$S_s((b_t)_{t\leq T})=\sum_{n=s}^{T}e^{\sum_{m=1}^{n}j_m(b_m)}=\sum_{n=s}^{T}P_n((b_t)_{t\leq T}),$$

for any $s\leq T$. For the rest of the proof, fix an argument sequence $(b_t)_{t\leq T}$. We will evaluate all the functions at this sequence, and consequently economize on notation by dropping the argument of S_s, P_s and j_s.

For unit vectors e_r and e_s, $r \geq s$, consider the derivatives of the function I:

$$\frac{\partial I}{\partial e_s} = j_s' S_s,$$
$$\frac{\partial^2 I}{\partial e_r \partial e_s} = j_s' j_r' S_r + \mathbf{1}_{r=s} j_s'' S_s.$$

Therefore, for a fixed vector $y = (y_t)_{t \leq T}$ the second derivative in the direction $y = (y_t)_{t \leq T}$ is:

$$\frac{\partial^2 I}{\partial y \partial y} = \sum_{s=1}^{T} \sum_{r=1}^{T} y_s y_r \frac{\partial^2 I}{\partial e_s \partial e_r} = 2 \sum_{s=1}^{T} \sum_{r \geq s} y_s y_r j_s' j_r' S_r + \sum_{s=1}^{T} y_s^2 j_s'' S_s. \tag{74}$$

We will bound the expression in (74). For this purpose note that for any $s \leq T$ and $q \leq T-s$ we have:

$$S_{s+q} = \sum_{n=s+q}^{T} P_n \leq e^{q \sup j_t} \sum_{n=s}^{T} P_n = e^{q \sup j_t} S_s,$$

It follows that for $\psi = \frac{\sup j_t}{2}$:

$$\sum_{r \geq s} S_r e^{-\psi(r-s)} \leq C S_s, \quad \sum_{s,r \geq s} S_r y_r^2 e^{\psi(r-s)} = \sum_{r} y_r^2 S_r \sum_{s \leq r} e^{\psi(r-s)} \leq C \sum_{s} S_s y_s^2, \tag{75}$$

where:

$$C = \sum_{n=0}^{T} e^{n\psi}. \tag{76}$$

Consequently, for any vector $z{=}(z_t)_{t \leq T}$, $z_t \in \mathbb{R}$:

$$\sum_{s,r \geq s} z_s z_r S_r = \sum_{s} z_s \sum_{r \geq s} \sqrt{S_r} z_r e^{\frac{\psi}{2}(r-s)} \sqrt{S_r} e^{-\frac{\psi}{2}(r-s)} \leq \sum_{s} z_s \left(\sum_{r \geq s} S_r z_r^2 e^{\psi(r-s)} \right)^{1/2} \left(\sum_{r \geq s} S_r e^{-\psi(r-s)} \right)^{1/2} \tag{77}$$
$$\leq \sqrt{C} \sum_{s} z_s \sqrt{S_s} \left(\sum_{r \geq s} S_r z_r^2 e^{\psi(r-s)} \right)^{1/2} \leq \sqrt{C} \left(\sum_{s} z_s^2 S_s \right)^{1/2} \left(\sum_{s} \left(\sum_{r \geq s} S_r z_r^2 e^{\psi(r-s)} \right) \right)^{1/2}$$
$$\leq C \left(\sum_{s} z_s^2 S_s \right)^{1/2} \left(\sum_{s} S_s z_s^2 \right)^{1/2} = C \sum_{s} z_s^2 S_s,$$

where the first and third inequalities follow from the Cauchy-Schwartz inequality, and C is as in (76).

Therefore, using both (74) and (77) we obtain:

$$\frac{\partial^2 I}{\partial y \partial y} \leq \sum_{s=1}^{T} y_s^2 \left(2C j_s'^2 + j_s'' \right) S_s,$$

establishing the Lemma. ■

Proof of Lemma 3 To show that $I\left((x_t)_{t\leq L}\right)$ is jointly concave in leisure $(x_t)_{t\leq L}$ we use Lemma 2 with $b_t = x_t$ and:

$$j_s(x_s) = \left(\theta_s - \phi\theta_{s+1}\right)\left(f(x_s) - a_s^*\right) + \ln\rho, \tag{78}$$

Since

$$f'(x_s) = \frac{-1}{g'(f(x_s))}, \quad f''(x_s) = \frac{-g''(f(x_s))}{g'^3(f(x_s))},$$

and we have assumed that $\theta_s - \phi\theta_{s+1} \geq 0$, the condition (52) is satisfied if g has sufficiently high curvature. ■

Proof of Lemma 4 We must verify condition (52) in Lemma 2 for $b_t = x_t$ and j_s defined as:

$$j_s(x_s) = \left(\theta_s - \phi\theta_{s+1}\right)\left(f(x_s) - \gamma a_s^*\right) + D_s,$$

for $D_s = (1-\gamma)k_s + \ln E\left(e^{(1-\gamma)\theta_s e_s}\right) + \ln\rho + (1-\gamma)\left(g\left(a_{s-1}^*\right) + g\left(a_s^*\right)\right)$. The rest of the proof follows as in the $\gamma = 1$ case, with the derivatives of the f function being:

$$f'(x_s) = -D\frac{1}{x_s g'(f(x_s))},\ f''(x_s) = \frac{1}{x_s^2 g'^2(f(x_s))}\left(Dg'(f(x_s)) - D^2\frac{g''(f(x_s))}{g'(f(x_s))}\right),$$

for $D = \frac{\gamma}{1-\gamma}sign(1-\gamma)$. Consequently $I'((x_t)_{t\leq L})$ is jointly concave. ■

Proof of Lemma 6 Let $U_t(\boldsymbol{\eta}_t;\eta_t')$ be the CEO's continuation utility after history $\boldsymbol{\eta}_t$ if the agent reports $\boldsymbol{\eta}_{t-1},\eta_t'$. (70) follows from the standard envelope conditions, i.e. $\frac{\partial}{\partial\eta_t'}U_t(\boldsymbol{\eta}_t;\eta_t')|_{\eta_t'=\eta_t} = 0$ together with:

$$U_t(\boldsymbol{\eta}_t;\eta_t') = U_t(\boldsymbol{\eta}_{t-1},\eta_t') + g(a_t(\boldsymbol{\eta}_{t-1},\eta_t')) - g(a_t(\boldsymbol{\eta}_{t-1},\eta_t') + \eta_t' - \eta_t),\ \text{for } \gamma = 1,$$

$$U_t(\boldsymbol{\eta}_t;\eta_t') = U_t(\boldsymbol{\eta}_{t-1},\eta_t') + \frac{y_t(\boldsymbol{\eta}_{t-1},\eta_t')^{1-\gamma}\left[e^{-g(a_t(\boldsymbol{\eta}_{t-1},\eta_t')+\eta_t'-\eta_t)(1-\gamma)} - e^{-g(a_t(\boldsymbol{\eta}_{t-1},\eta_t'))(1-\gamma)}\right]}{1-\gamma}.\ \text{for } \gamma \neq 1.$$

The technical assumptions on $a_t(\boldsymbol{\eta}_{t-1},\cdot)$ guarantee that $U_t(\boldsymbol{\eta}_{t-1},\cdot)$ is absolutely continuous (see EG for details). $y_t(\boldsymbol{\eta}_t) > 0$ follows from PS, since the marginal utility of consumption at zero is infinite. ■

Proof of Lemma 7 Note that if instead of $U_t^{\#}(\boldsymbol{\eta}_{t-1},\cdot)$ and $\zeta(\boldsymbol{\eta}_{t-1},\cdot)$ we solve for the functions $\overline{U_t^{\#}}(\boldsymbol{\eta}_{t-1},\cdot)$ and $\overline{\zeta}(\boldsymbol{\eta}_{t-1},\cdot)$ that satisfy $\overline{U_t^{\#}}(\boldsymbol{\eta}_{t-1},\underline{\eta}) = U_t(\boldsymbol{\eta}_{t-1},\underline{\eta})$ and

$$\overline{U_t^{\#}}(\boldsymbol{\eta}_{t-1},\eta_t) = \overline{U_t^{\#}}(\boldsymbol{\eta}_{t-1},\underline{\eta}) + \int_{\underline{\eta}}^{\eta_t}[\overline{\zeta}(\boldsymbol{\eta}_{t-1},x)y_t(\boldsymbol{\eta}_{t-1},x)e^{-g(\overline{a})}]^{1-\gamma}g'(\overline{a})dx, \tag{79}$$

$$\overline{U_t^{\#}}(\boldsymbol{\eta}_{t-1},\eta_t) - U_t(\boldsymbol{\eta}_{t-1},\eta_t) = g(a_t(\boldsymbol{\eta}_{t-1},\eta_t)) - g(\overline{a}) + \ln\overline{\zeta}(\boldsymbol{\eta}_{t-1},\eta_t), for \gamma = 1.$$

$$\frac{\overline{U_t^{\#}}(\boldsymbol{\eta}_{t-1},\eta_t)}{U_t(\boldsymbol{\eta}_{t-1},\eta_t)} = \frac{[\overline{\zeta}(\boldsymbol{\eta}_{t-1},\eta_t)y_t(\boldsymbol{\eta}_{t-1},\eta_t)e^{-g(\overline{a})}]^{1-\gamma}}{[y_t(\boldsymbol{\eta}_{t-1},\eta_t)e^{-g(a_t(\boldsymbol{\eta}_{t-1},\eta_t))}]^{1-\gamma}},\ \text{for } \gamma \neq 1,$$

then we have $\zeta(\boldsymbol{\eta}_{t-1}, \eta_t) \leq \overline{\zeta}(\boldsymbol{\eta}_{t-1}, \eta_t)$ (and $\zeta(\boldsymbol{\eta}_{t-1}, \eta_t) = \overline{\zeta}(\boldsymbol{\eta}_{t-1}, \eta_t)$ when $t = L$). Therefore it will be sufficient to $E_{t-1}\left[\overline{\zeta}(\boldsymbol{\eta}_{t-1}, \eta_t)\right]$.

Since $\boldsymbol{\eta}_{t-1}$ is fixed, to economize on notation we write $U_t(\eta_t)$ instead of $U_t(\boldsymbol{\eta}_{t-1}, \eta_t)$ etc.

Case $\gamma \neq 1$. We have:

$$\overline{U_t^{\#}}(\eta_t) = \overline{U_t^{\#}}(\underline{\eta}) + \int_{\underline{\eta}}^{\eta_t} \frac{\overline{U_t^{\#}}(x)}{U_t(x)} \left[y_t(x)e^{-g(a_t(x))}\right]^{1-\gamma} g'(a_t(x)) \frac{g'(\overline{a})}{g'(a_t(x))} dx,$$
$$U_t(\eta_t) = \overline{U_t^{\#}}(\underline{\eta}) + \int_{\underline{\eta}}^{\eta_t} \left[y_t(x)e^{-g(a_t(x))}\right]^{1-\gamma} g'(a_t(x)) dx.$$

Therefore:

$$\left(\frac{\overline{U_t^{\#}}(\eta_t)}{U_t(\eta_t)}\right)' =$$
$$= \frac{\frac{\overline{U_t^{\#}}(\eta_t)}{U_t(\eta_t)} \left[y_t(\eta_t)e^{-g(a_t(\eta_t))}\right]^{1-\gamma} g'(a_t(\eta_t)) \frac{g'(\overline{a})}{g'(a_t(\eta_t))} U_t(\eta_t) - \left[y_t(\eta_t)e^{-g(a_t(\eta_t))}\right]^{1-\gamma} g'(a_t(\eta_t)) \overline{U_t^{\#}}(\eta_t)}{U_t(\eta_t)^2} =$$
$$= \frac{\overline{U_t^{\#}}(\eta_t) \left[y_t(\eta_t)e^{-g(a_t(\eta_t))}\right]^{1-\gamma} g'(a_t(\eta_t)) \left[\frac{g'(\overline{a})}{g'(a_t(\eta_t))} - 1\right]}{U_t(\eta_t)^2} \leq \frac{\overline{U_t^{\#}}(\eta_t)}{U_t(\eta_t)} (1-\gamma) g'(\overline{a}) \left[\frac{g'(\overline{a})}{g'(a_t(\eta_t))} - 1\right] \text{ for } \gamma < 1,$$

It follows that:

$$\frac{\overline{U_t^{\#}}(\eta_t)}{U_t(\eta_t)} \leq e^{(1-\gamma) g'(\overline{a}) \int_{\underline{\eta}}^{\eta_t} \left(\frac{g'(\overline{a})}{g'(a_t(x))} - 1\right) dx} \leq e^{(1-\gamma) \sup \frac{g'(\overline{a})}{f} E_{t-1}\left(\frac{g'(\overline{a})}{g'(a_t(x))} - 1\right)} \leq e^{(1-\gamma) g'(\overline{a}) \sup \frac{g''}{f g'^2} E_{t-1}[\overline{a} - a_t(\boldsymbol{\eta}_t)]}, \text{ for } \gamma < 1. \tag{80}$$

where the last inequality follows because $\frac{g'(\overline{a})}{g'(a)} = g'(\overline{a}) \left[\frac{1}{g'(\overline{a})} + (\overline{a} - a) \frac{g''(x\overline{a} + (1-x)a)}{g'^2(x\overline{a} + (1-x)a)}\right]$ for some $x \in [0, 1]$. For $\gamma > 1$ we obtain the analogous chain with the inequality signs reversed. Thus,

$$E_{t-1}\left[\overline{\zeta}(\eta_t)\right] = E_{t-1}\left[\left[\frac{\overline{U_t^{\#}}(\eta_t)}{U_t(\eta_t)}\right]^{\frac{1}{1-\gamma}} e^{[g(\overline{a}) - g(a_t(\eta_t))](1-\gamma)}\right] \leq \tag{81}$$
$$\leq e^{g'(\overline{a}) \sup \frac{g''}{f g'^2} E_{t-1}[\overline{a} - a_t(\eta_t)]} E_{t-1}\left[e^{[g(\overline{a}) - g(a_t(\eta_t))](1-\gamma)}\right] \leq$$
$$\leq e^{g'(\overline{a}) \sup \frac{g''}{f g'^2} E_{t-1}[\overline{a} - a_t(\eta_t)]} \left(1 + \mathbf{1}_{\gamma < 1} e^{g(\overline{a}) - g(\underline{a})} (1-\gamma) g'(\overline{a}) E_{t-1}\left[\overline{a} - a_t(\eta_t)\right]\right).$$

Case $\gamma = 1$. Comparing (70) and (79) we immediately obtain:

$$\ln \overline{\zeta}(\eta_t) = \int_{\underline{\eta}}^{\eta_t} \left(\frac{g'(\overline{a})}{g'(a_t(x))} - 1\right) g'(a_t(x)) dx + g(\overline{a}) - g(a_t(\eta_t)).$$

Using the analogous bounds as in (80) and (81) we obtain:

$$E_{t-1}\left[\overline{\zeta}(\eta_t)\right] \leq E_{t-1}\left[e^{g'(\overline{a})\int_{\underline{\eta}}^{\eta_t}\left(\frac{g'(\overline{a})}{g'(a_t(x))}-1\right)dx+g(\overline{a})-g(a_t(x))}\right] \leq e^{g'(\overline{a})\sup\frac{g''}{fg'^2}E_{t-1}[\overline{a}-a_t(\eta_t)]}E_{t-1}\left[e^{g(\overline{a})-g(a_t(\eta_t))}\right] \leq$$
$$\leq e^{g'(\overline{a})\sup\frac{g''}{fg'^2}E_{t-1}[\overline{a}-a_t(\eta_t)]}\left(1+e^{g(\overline{a})-g(\underline{a})}g'(\overline{a})E_{t-1}\left[\overline{a}-a_t(\eta_t)\right]\right).$$

■

Proof of Lemma 8 Multiplying all payoffs by ζ results in all the continuation utilities $U_t(\boldsymbol{\eta}_t)$ and deviation continuation utilities $U_t(\boldsymbol{\eta}_t;\eta_t')$ multiplied by constant $\zeta^{1-\gamma}$ for $\gamma\neq 1$, or having a constant $\ln\zeta\times\sum_{s=0}^{L-t}\rho^s$ added at time t, for $\gamma=1$, and so EF is unaffected. This also results in the marginal utilities of current consumption multiplied by $\zeta^{-\gamma}$, and so PS is also unaffected.

■

Proof of Lemma 9 We prove only the $\gamma\neq 1$ case. For the $\overline{\zeta}$ as in the proof of Lemma (7) we have:

$$E_{t-1}\left[\frac{mu_t^{\#}(\boldsymbol{\eta}_{t-1},\eta_t)}{mu_t(\boldsymbol{\eta}_{t-1},\eta_t)}\right] \geq E_{t-1}\left[\overline{\zeta}^{-\gamma}(\boldsymbol{\eta}_{t-1},\eta_{t-1})\times e^{(1-\gamma)(g(a(\boldsymbol{\eta}_{t-1},\eta_t))-g(\overline{a}))}\right] =$$
$$= E_{t-1}\left[\left[\frac{\overline{U_t^{\#}}(\boldsymbol{\eta}_{t-1},\eta_t)}{U_t(\boldsymbol{\eta}_{t-1},\eta_t)}\right]^{\frac{-\gamma}{1-\gamma}} e^{-\gamma(1-\gamma)[g(\overline{a})-g(a_t(\boldsymbol{\eta}_{t-1},\eta_t))]}\times e^{(1-\gamma)(g(a_t(\boldsymbol{\eta}_{t-1},\eta_t))-g(\overline{a}))}\right] =$$
$$\geq e^{-\gamma g'(\overline{a})\sup\frac{g''}{fg'^2}E_{t-1}[\overline{a}-a_t(\boldsymbol{\eta}_t)]}E_{t-1}\left[e^{-(1+\gamma)(1-\gamma)[g(\overline{a})-g(a_t(\boldsymbol{\eta}_{t-1},\eta_t))]}\right] \geq$$
$$\geq e^{-\gamma g'(\overline{a})\sup\frac{g''}{fg'^2}E_{t-1}[\overline{a}-a_t(\boldsymbol{\eta}_t)]}\left(1-\mathbf{1}_{\gamma<1}e^{-(1+\gamma)(1-\gamma)[g(\overline{a})-g(\underline{a})]}g'(\overline{a})(1-\gamma)(1+\gamma)E_{t-1}\left[\overline{a}-a_t(\boldsymbol{\eta}_t)\right]\right).$$

■

Proof of Lemma 10 We prove only the $\gamma\neq 1$ case. From (70) it follows that for every η_t and η_t' :

$$e^{(\overline{\eta}-\underline{\eta})g'(\overline{a})}\times y_t^h(\boldsymbol{\eta}_{t-1},\eta_t)^{1-\gamma}e^{-(1-\gamma)g(a_t^h(\boldsymbol{\eta}_{t-1},\eta_t))} \geq y_t^h(\boldsymbol{\eta}_{t-1},\eta_t')^{1-\gamma}e^{-(1-\gamma)g(a_t^h(\boldsymbol{\eta}_{t-1},\eta_t'))},$$

and so for every η_t and η_t' :

$$y^h(\boldsymbol{\eta}_{t-1},\eta_t')^{-\gamma}e^{\gamma g(a_t^h(\boldsymbol{\eta}_{t-1},\eta_t'))} \geq e^{-\left|\frac{\gamma}{1-\gamma}\right|(\overline{\eta}-\underline{\eta})g'(\overline{a})}\times y_t^h(\boldsymbol{\eta}_{t-1},\eta_t)^{-\gamma}e^{\gamma g(a^h(\boldsymbol{\eta}_{t-1},\eta_t))},$$
$$E_{t-1}\left[mu_t^h(\boldsymbol{\eta}_{t-1},\eta_t)\right] \geq e^{-\left|\frac{\gamma}{1-\gamma}\right|(\overline{\eta}-\underline{\eta})g'(\overline{a})+g(\underline{a})-g(\overline{a})}\times\max_x mu_t^h(\boldsymbol{\eta}_{t-1},x).$$

It follows that for $D_2 = e^{\left|\frac{\gamma}{1-\gamma}\right|(\overline{\eta}-\underline{\eta})g'(\overline{a})+g(\overline{a})-g(\underline{a})}$,

$$\frac{E_{t-1}\left[mu_t^l(\boldsymbol{\eta}_{t-1},\eta_t)\right]}{E_{t-1}\left[mu_t^h(\boldsymbol{\eta}_{t-1},\eta_t)\right]} \geq \frac{E_{t-1}\left[mu_t^h(\boldsymbol{\eta}_{t-1},\eta_t)\right]\left(1-D_2\times\left(1-E_{t-1}\left[\frac{mu_t^l(\boldsymbol{\eta}_{t-1},\eta_t)}{mu_t^h(\boldsymbol{\eta}_{t-1},\eta_t)}\right]\right)\right)}{E_{t-1}\left[mu_t^h(\boldsymbol{\eta}_{t-1},\eta_t)\right]} =$$
$$= 1 - D_2 \times \left(1 - E_{t-1}\left[\frac{mu_t^l(\boldsymbol{\eta}_{t-1},\eta_t)}{mu_t^h(\boldsymbol{\eta}_{t-1},\eta_t)}\right]\right).$$

■

Proof of Lemma 11 Let Y^0 be the payoff scheme Y^x. For any n, $0 < n < L$, we construct the payoff scheme Y^n as follows. Start with the payoff scheme Y^{n-1}. After any history $\boldsymbol{\eta}_n$ multiply the payoffs at time n by $\zeta^{n,ps}(\boldsymbol{\eta}_n) > 1$ so that PS at history $\boldsymbol{\eta}_n$ is satisfied; then multiply the payoffs after any history $\boldsymbol{\eta}_m$, $m \geq n$ and $\boldsymbol{\eta}_{m|n} = \boldsymbol{\eta}_n$, by $\zeta^{n,pu}(\boldsymbol{\eta}_n) < 1$ so that the continuation utility at history $\boldsymbol{\eta}_n$ remains unchanged. After any history $\boldsymbol{\eta}_{n-1}$ multiply the payoffs at time $n-1$ by $\zeta^{n,ps}(\boldsymbol{\eta}_{n-1}) > 1$ so that PS at $\boldsymbol{\eta}_{n-1}$ is satisfied; then multiply the payoffs after any history $\boldsymbol{\eta}_m$, $m \geq n-1$ and $\boldsymbol{\eta}_{m|n-1} = \boldsymbol{\eta}_{n-1}$, by $\zeta^{n,pu}(\boldsymbol{\eta}_{n-1}) < 1$ so that the continuation utility at $\boldsymbol{\eta}_{n-1}$ remains unchanged. Follow this procedure until histories at time 1, and let Y^n be the resulting payoff scheme. One can inductively show that $\zeta^{n,pu}(\boldsymbol{\eta}_m)\times\,\zeta^{n,ps}(\boldsymbol{\eta}_m) \geq 1$, $m \leq n$.

Let A^* always require the high effort. Lemma 8 yields that each contract (A^*, Y^n) satisfies EF and also PS up to round n. Let $Y^* = Y^{L-1}$. It remains to prove (73).

For any history $\boldsymbol{\eta}_L$ we have $y_L^*(\eta_L) = y_L^x(\eta_L) \times \prod_{m=1}^{L-1}\prod_{n=m}^{L-1}\zeta^{n,pu}(\boldsymbol{\eta}_{L|m}) \leq y_L^x(\eta_L)$ and so the condition (73) is satisfied.

For any history $\boldsymbol{\eta}_t$, $t < L$, we have, by construction above:

$$\frac{mu_t^*(\boldsymbol{\eta}_t)}{mu_t^x(\boldsymbol{\eta}_t)} = \left(\prod_{m=1}^{t}\prod_{n=m}^{L-1}\zeta^{n,pu}(\boldsymbol{\eta}_{t|m}) \times \prod_{n=t}^{L-1}\zeta^{n,ps}(\boldsymbol{\eta}_t)\right)^{-\gamma} \geq \left(\prod_{n=t}^{L-1}\zeta^{n,ps}(\boldsymbol{\eta}_t)\right)^{-\gamma}.$$

Moreover,

$$\zeta^{t,ps}(\boldsymbol{\eta}_t)^{-\gamma} = \frac{E_t\left[mu_{t+1}^x(\boldsymbol{\eta}_{t-1},\eta_t)\right]}{E_t\left[mu_{t+1}(\boldsymbol{\eta}_{t-1},\eta_t)\right]} \geq \phi\left(E_t\left[\frac{mu_{t+1}^x(\boldsymbol{\eta}_{t-1},\eta_t)}{mu_{t+1}(\boldsymbol{\eta}_{t-1},\eta_t)}\right]\right) \geq$$
$$\geq \phi\left(\psi\left(E_t\left[\overline{a} - a_{t+1}(\eta_{t+1})\right]\right)\right),$$

where the first inequality follows from Lemma 10, and the second one from Lemma 9. By the same logic, for any n, $t < n \leq L-1$,

$$\begin{aligned}
\zeta^{n,ps}(\boldsymbol{\eta}_t)^{-\gamma} &= \frac{E_t\left[mu_{t+1}^{n}(\boldsymbol{\eta}_t,\eta_{t+1})\right]}{E_t\left[mu_{t+1}^{n-1}(\boldsymbol{\eta}_t,\eta_{t+1})\right]} \geq \phi\left(E_t\left[\frac{mu_{t+1}^{n}(\boldsymbol{\eta}_t,\eta_{t+1})}{mu_{t+1}^{n-1}(\boldsymbol{\eta}_t,\eta_{t+1})}\right]\right) \geq \phi\left(E_t\left[\zeta^{n,ps}(\boldsymbol{\eta}_t,\eta_{t+1})^{-\gamma}\right]\right) \\
&= \phi\left(E_t\left[\frac{E_{t+1}\left[mu_{t+2}^{n}(\eta_t,\eta_{t+1},\eta_{t+2})\right]}{E_{t+1}\left[mu_{t+2}^{n-1}(\boldsymbol{\eta}_t,\eta_{t+1},\eta_{t+2})\right]}\right]\right) \geq \phi\left(E_t\left[\phi\left(E_{t+1}\left[\frac{mu_{t+2}^{n}(\boldsymbol{\eta}_t,\eta_{t+1},\eta_{t+2})}{mu_{t+2}^{n-1}(\boldsymbol{\eta}_t,\eta_{t+1},\eta_{t+2})}\right]\right)\right]\right) \\
&= \phi^2\left(E_t\left[\frac{mu_{t+2}^{n}(\boldsymbol{\eta}_t,\eta_{t+1},\eta_{t+2})}{mu_{t+2}^{n-1}(\boldsymbol{\eta}_t,\eta_{t+1},\eta_{t+2})}\right]\right) \geq ... \geq \phi^{n-t}\left(E_t\left[\frac{mu_{n}^{n}(\boldsymbol{\eta}_t,\eta_{t+1},...,\eta_n)}{mu_{n}^{n-1}(\boldsymbol{\eta}_t,\eta_{t+1},...,\eta_n)}\right]\right) \\
&\geq \phi^{n-t}\left(E_t\left[\zeta^{n,ps}(\boldsymbol{\eta}_t,\eta_{t+1},...,\eta_n)^{-\gamma}\right]\right) = \phi^{n-t}\left(E_t\left[\frac{E_n\left[mu_{n+1}^{x}(\boldsymbol{\eta}_t,\eta_{t+1},...,\eta_{n+1})\right]}{E_n\left[mu_{n+1}(\boldsymbol{\eta}_t,\eta_{t+1},...,\eta_{n+1})\right]}\right]\right) \\
&\geq \phi^{n-t+1}\left(E_t\left[\psi\left(E_n\left[\overline{a}-a_{n+1}(\boldsymbol{\eta}_t,\eta_{t+1},...,\eta_{n+1})\right]\right)\right]\right).
\end{aligned}$$

■